Biost 517
Applied Biostatistics I

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Review

December 6, 2010

1

© 2002, 2003, 2005 Scott S. Emerson, M.D., Ph.D.

Overview of Setting

Scientific Method

2

Purpose of Statistics

- Statistics is about science
  - (Science in the broadest sense of the word)
- Science is about proving things to people
  - (The validity of any proof rests solely on the willingness of the audience to believe it)

3

First Stage of Scientific Investigation

- Hypothesis generation
  - Observation
    - Measurement of existing populations
    - Disadvantages:
      - Confounding
      - Limited ability to establish cause and effect

4

Further Stages of Scientific Investigation

- Refinement and confirmation of hypotheses
  - Experiment
    - Intervention
    - Elements of experiment
      - Overall goal
      - Specific aims (hypotheses)
      - Materials and methods
      - Collection of data
      - Analysis
      - Interpretation; Refinement of hypotheses

5

Scientific Method: Key Elements

- Overall goal
- Specific aims (hypotheses)
- Materials and methods
- Collection of data
- Analysis
- Interpretation

6

Do You Need Statistics?

- Two question test (Both must be YES)
  - In a deterministic world, do YOU know how to answer your question?
    - Is the question answerable in the real world?
    - How do you use a number to answer the scientific question?
  - In a world subject to variation, do YOU know how you would answer your question if you had the entire population?

7

Statistical Tasks

- Understand overall goal
- Refine specific aims (stat hypotheses)
- Materials and methods: Study design
- Collection of data: Advise on QC
- Analysis
  - Describe sample (materials and methods)
  - Analyses to address specific aims
- Interpretation

8

Statistical Tasks

Statistical Hypotheses

9

Statistical Questions

- Clustering of observations
- Clustering of variables
- Quantification of distributions
- Comparing distributions
- Prediction of individual observations

10

Scientific Hypotheses:

- Usual statement:
  - The intervention when given to the target population will tend to result in outcome measurements that are

higher than,
lower than, or
about the same as

an absolute standard, or
measurements in a comparison group

11

Statistics and Game Theory

- Multiple comparison issues
  - Different clinical endpoints, summary measures, modeling of predictors, adjustment for covariates, subgroups
  - Type I error for each analysis
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025
- Multiple endpoints increase the chance of deciding an ineffective treatment should be adopted
  - This problem exists with either frequentist or Bayesian criteria for evidence
  - The actual inflation of the type I error depends
    - the number of multiple comparisons, and
    - the correlation between the endpoints

12

Ex: Level 0.05 per Decision

- Experiment-wise Error Rate

Number Compared	Worst Case	Correlation 0.00	0.30	0.50	0.75	0.90
1	.050	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081	.070
3	.150	.143	.137	.126	.104	.084
5	.250	.226	.208	.184	.138	.101
10	.500	.401	.353	.284	.193	.127
20	1.000	.642	.540	.420	.258	.154
50	1.000	.923	.806	.624	.353	.193

13

Statistical Role of Variables

- Statistical hypotheses involve
  - "Response" or "Outcome"
    - Can be either the "effect" or the "cause"
  - "Grouping Variable(s)"
    - Primary scientific question
      - Predictor of interest
      - Effect Modifiers
    - Adjustment for covariates
      - Confounders
      - Precision variables

14

An Aside:

Ability to

Detect Associations

15

Definition of an Association

- The distributions of two variables are not independent
- Independence: Equivalent definitions
  - Probability of outcome and exposure is product of
    - Overall probability of outcome, and
    - Overall probability of exposure
  - Distribution of exposure is the same across all outcome categories
  - Distribution of outcome is the same across all exposure categories

16

Mathematical Definitions

- Independence: Equivalent definitions
  - Joint probability of outcome O and cause C
    - $\Pr(O = o_1, C = c_1) = \Pr(O = o_1) \times \Pr(C = c_1)$
  - Conditional probability of outcome given cause
    - $\Pr(O = o_1 \mid C = c_1) = \Pr(O = o_1 \mid C = c_2)$
  - Conditional probability of cause given outcome
    - $\Pr(C = c_1 \mid O = o_1) = \Pr(C = c_1 \mid O = o_2)$

17

Establishing Independence

- Consider all events defined by the two variables
- For each choice of o_1, o_2, c_1, c_2 show either
  - $\Pr(O = o_1, C = c_1) = \Pr(O = o_1) \times \Pr(C = c_1)$,
  - $\Pr(O = o_1 \mid C = c_1) = \Pr(O = o_1 \mid C = c_2)$, or
  - $\Pr(C = c_1 \mid O = o_1) = \Pr(C = c_1 \mid O = o_2)$

18

Detecting Associations

- It takes an infinite sample size to prove equality
- Instead, we detect associations by showing that two variables are not independent
  - Thus, we show that two distributions are different

19

Summary Measures

- Generally we consider some summary measure of the distribution
- E.g., when we use the mean, we show an association by showing either
  - $E(O \times C) \neq E(O) \times E(C)$,
  - $E(O \mid C = c_1) \neq E(O \mid C = c_2)$, or
  - $E(C \mid O = o_1) \neq E(C \mid O = o_2)$

20

Justification

- This works, because if two distributions are the same, ALL summary measures should be the same
- If some summary measure is different, then we know the distributions are different

21

Hierarchy of Null Hypotheses

- Strong Null
  - Distribution of response identical in all groups
- Intermediate Null
  - Summary measure identical in all groups
    - Summary measures on a flat line
- Weak Null
  - No linear trend in summary measure across groups
    - On average, summary measures on a flat line

22

Impact of Study Design

- Study design affects the way we can establish an association
- Cohort studies fix the sample size in exposure groups
  - They must examine whether
    - $Pr(O \mid C = c_1) \neq Pr(O \mid C = c_2)$
- Case-control studies fix the sample size in outcome groups
  - They must examine whether
    - $Pr(C \mid O = o_1) \neq Pr(C \mid O = o_2)$
- Cross sectional studies only fix the total sample size
  - They can examine either of the above, as well as whether
    - $Pr(O, C) \neq Pr(O) \times Pr(C)$

23

Summary Measures

24

For Each Outcome Define "Tends To"

- In general, the space of all probability distributions is not totally ordered
- There are an infinite number of ways we can define a tendency toward a "larger" outcome
- This can be difficult to decide even when we have data on the entire population
  - Ex: Is the highest paid occupation in the US the one with
    - the higher mean?
    - the higher median?
    - the higher maximum?
    - the higher proportion making $1M per year?

25

Statistical Issues

- Need to choose a primary summary measure or multiple comparison issues result
- Example: Type I error with normal data
  - Any single test: 0.050
  - Mean, geometric mean 0.057
  - Mean, Wilcoxon 0.061
  - Mean, geom mean, Wilcoxon 0.066
  - Above plus median 0.085
  - Above plus Pr (Y > 1 sd) 0.127
  - Above plus Pr (Y > 1.645 sd) 0.169

26

Univariate Summary Measures

- Many times, statistical hypotheses are stated in terms of summary measures for the distribution within groups
  - Means (arithmetic, geometric, harmonic, …)
  - Medians (or other quantiles)
  - Proportion exceeding some threshold
  - Odds of exceeding some threshold
  - Time averaged hazard function (instantaneous risk)
  - …

27

Comparisons Across Groups

- Comparisons across groups then use differences or ratios
  - Difference / ratio of means (arithmetic, geometric, …)
  - Difference / ratio of proportion exceeding some threshold
  - Difference / ratio of medians (or other quantiles)
  - Ratio of odds of exceeding some threshold
  - Ratio of hazard (averaged across time?)
  - …

28

Based on Type of Data

- Correspondence to relevance of descriptive statistics
  - Binary or dichotomous:
    - mean (proportion); odds
  - Nominal (unordered categories):
    - frequencies; odds
  - Ordinal (ordered categories):
    - median (quantiles); odds; ? mean
  - Quantitative (addition makes sense):
    - mean; median; proportion > c; hazards, …

29

Univariate Descriptive Statistics

		Binary	Unordered Nominal	Ordered Categ	Ordered Quant	Ordered Cens
Entire Distnution	**Frequency**	OK	OK	OK	OK	
	Cum Freq	**boring**		OK	OK	KM
	Mode	**boring**	Sample	Sample	Density	
	Min / Max	**boring**		**boring**	OK	
Dicho-tomize	**Proportion (or Odds)**	OK	OK	OK	OK	KM
Quant-iles	**Quantiles (25^{th},Mdn,75^{th})**	**boring**		OK	OK	KM
Means	**Arithmetic**	(Prop)		***	OK	(?KM)
	Geometric				OK	(?KM)
	Harmonic				OK	(?KM)
	Std Dev	**boring**			OK	(?KM)
	Skew, Kurt	**boring**			OK	(?KM)

Commonly Used Parameters

		Binary	Unordered Nominal	Ordered Categ	Ordered Quant	Ordered Cens
Entire Distnution	**Frequency**	(Prop)	OK	OK	(categ)	
	CDF / Surv			OK	OK	OK
	Mode					
	Min / Max				(Difficult)	
Dicho-tomize	**Proportion (or Odds)**	OK	OK	OK	OK	OK
Quant-iles	**Quantiles (25^{th},Mdn,75^{th})**			OK	OK	OK
Means	**Arithmetic**	(Prop)		(OK)	OK	(Restr)
	Geometric				OK	(Restr)
	Harmonic				OK	(Restr)
	Std Dev					
	Skew, Kurt					

Joint Summary Measures

- Other times groups are compared using a summary measure for the joint distribution
  - Median difference / ratio of paired observations
  - Probability that a randomly chosen measurement from one population might exceed that from the other
  - Maximal difference between cumulative distribution functions
  - …

32

Criteria for Summary Measure

- In order of importance
  - Scientifically (clinically) relevant
    - Also reflects current state of knowledge
  - Is likely to vary across levels of the factor of interest
    - Ability to detect variety of changes
  - Statistical precision
    - Only relevant if all other things are equal

33

Science vs Statistics

- Scientific summary measures
  - Summarize distributions of meaningful measurements
  - Contrasts across populations
    - E.g., a slope
- Statistical measures
  - How precisely we estimate a scientific measure
    - E.g., a P value, correlation

34

Statistical Tasks

Data Analysis

35

Before Looking At Data

- Overall goal
- Refine to specific aims
- Scientific classification of variables
  - Demographic, severity of disease, etc.
- Statistical classification of variables
  - Response
  - Predictor(s) of interest
  - Effect modifiers
  - (Potential) confounders
  - Precision variables

36

Effect Modifier

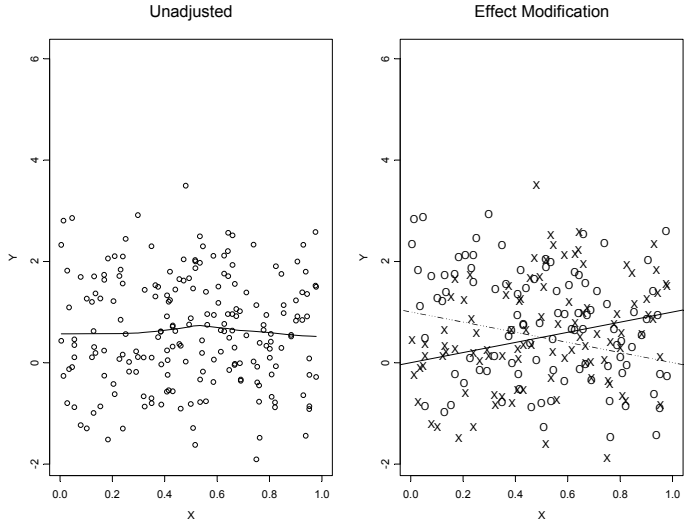
- The association between Response and POI differs in strata defined by effect modifier
  - Statistical term: “Interaction”
- Depends on the measurement of effect
  - Summary measure
    - Mean, geometric mean, median, proportion, odds, hazard, etc.
  - Comparison across groups
    - Difference, ratio

37

Effect modifiers: Graphs

- Estimates of treatment effect differ among the strata
- When analyzing difference of means of continuous data
  - Stratified smooth curves of data are nonparallel
- (Graphical techniques difficult in other settings)

38



Analysis of Effect Modification

- When the scientific question involves effect modification, analyses must be within each stratum separately
- If we want to estimate degree of effect modification or test for its existence:
  - A regression model will typically include
    - Predictor of interest (main effect)
    - Effect modifying variable (main effect)
    - A covariate modeling the interaction (usually product)

40

Ignoring Effect Modification

- By design or mistake, we sometimes do not model effect modification
- We might perform either
  - Unadjusted analysis:
    - POI only
  - Adjusted analysis:
    - POI and third variable, but no interaction term

41

Confounding

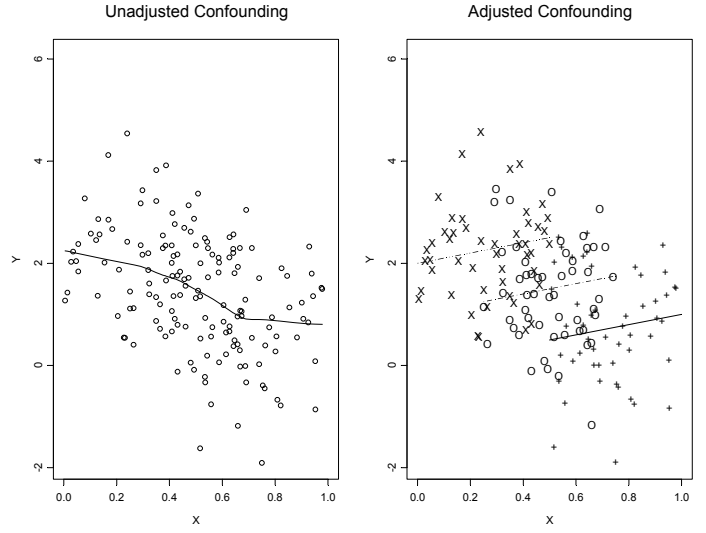
- The association between a predictor of interest and the response variable is confounded by a third variable if
  - The third variable is associated with the predictor of interest in the sample, AND
  - The third variable is associated with the response
    - causally (in truth)
    - in groups that are homogeneous with respect to the predictor of interest, and
    - not in the causal pathway of interest

42

Confounders: Graphs

- Estimates of treatment effect the same across strata, AND
  - Confounder is causally associated with Response independent of POI, AND
  - Confounder associated with POI in the sample
- When analyzing difference of means of continuous data
  - Stratified smooth curves of data are parallel
  - Distribution of POI differs across strata
  - Unadjusted, adjusted analyses give different estimates

43



Adjusting for confounding

- Based on beliefs about the causal relationships among measured variables
  - We cannot assess causal relationships in our statistical analysis
    - Causation inference comes only from study design
- Based on possibility of associations in our sample
  - Assoc between POI and confounder need not exist in population

45

Testing for confounding

- WE CANNOT DO THIS
- Confounding is product of
  - Association between POI, third variable in sample
    - P value is about associations in population
    - Small association with strong predictor ➔ confounding
  - Causal association between third variable / response independent of POI
    - Small causal effect with strong POI, third variable association in sample ➔ confounding
    - Confounding may also diminish ability to assess association in adjusted analysis

46

Unadjusted, Adjusted Analyses

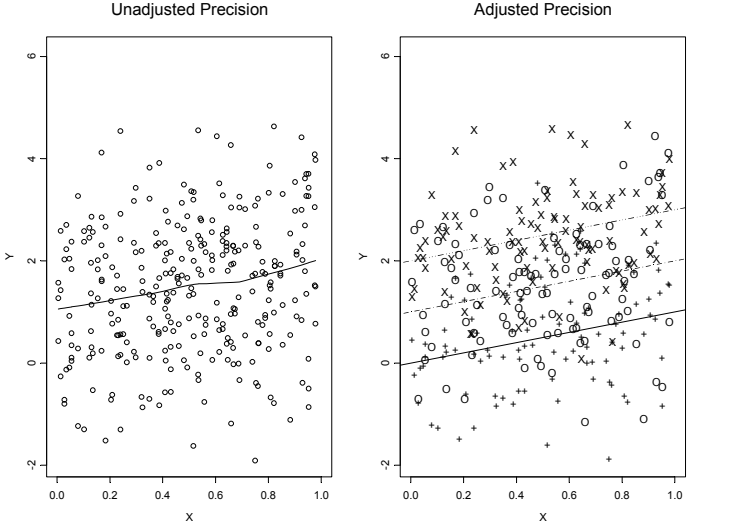
- Confounding typically produces a difference between unadjusted and adjusted analyses, but those symptoms are not proof of confounding
- Such a difference can occur times when there is no confounding

47

Precision Variables

- Estimates of treatment effect the same across strata, AND
  - Variable is causally associated with Response, AND
  - Variable not associated with POI in the sample
- When analyzing difference of means of continuous data
  - Stratified smooth curves of data are parallel
  - Distribution of POI same across strata
  - Unadjusted, adjusted analyses give similar estimates

48



Descriptive Statistics

- Identify measurement or data entry errors
  - Missing data, min, maxi
- Characterize materials and methods
  - Mean, SD, min, max, median, quartiles
- Validity of scientific,statistical assumptions
  - Confounding (?means), linearity (modeled summary measure), homoscedasticity (SD)
- (Simple estimates of effects-- inference)
- Hypothesis generation

50

Inference

- Generalizations from sample to population
  - Estimation
    - Point estimates
    - Interval estimates
  - Decision analysis (testing)
    - Quantifying strength of evidence

51

An Aside: Reporting Associations

- Hypothetical study to detect an association between Event B and Exposure F
  - Unexposed: 0 of 5 have Event B
    - Estimated incidence rate: 0.000
    - 95% CI for incidence rate: 0.000 – 0.522
  - Exposed: 3 of 5 have Event B
    - Estimated incidence rate: 0.600
    - 95% CI for incidence rate: 0.147 – 0.947
  - Fisher's Exact two-sided P: 0.167
- How would you characterize the presence of an association between these two variables?

52

WRONG Criteria

- Incorrect criteria for stating the existence of a statistically significant association
  - "Because the confidence intervals overlap, there is no association."
  - (We need to use a P value. The use of confidence intervals in this manner is more complicated.)

53

Independent CI and Tests

- Rules for **independent** strata
  - IF two independent 95% CI do not overlap
    - THEN we know a statistically significant difference exists (? P less than .006?)
  - IF the 95% CI for one stratum contains the point estimate of the other stratum
    - THEN we know the difference is not statistically significant (? P greater than .16?)
  - OTHERWISE all bets are off
    - Especially: we cannot reverse the above claims

54

WRONG

- An overstated, purely statistical report
  - "As the P value is greater than 0.05, we conclude that there is no association between exposure F and event B."
  - (We should not conclude that there is no association, because we lacked precision to rule out differences that might be of interest.)

55

Scientifically USELESS

- A correctly stated, purely statistical report
  - "As the P value is greater than 0.05, we conclude that there is not sufficient evidence to rule out the possibility of no association between exposure F and event B."
  - (Stated correctly, but gives no idea of whether we had ruled out differences that we cared about or we had merely done an abysmal study.)

56

CORRECT and USEFUL

- Scientific estimates and quantification of statistical evidence
  - "Incidence rates of 60% in the exposed (95% CI: 15% - 95%) and 0% in the unexposed (95% CI: 0% - 52%). Unfortunately, the precision was not adequate to demonstrate that such a large difference in incidence rates would be unlikely in the absence of a true association (P = 0.17)."
  - (These data are not atypical of setting in which F= female and B= giving birth.)

57

Statistical Tasks

Analysis Methods

58

Biost 517

- We described tests (and sometimes CI) for comparing parameters across groups
- Not all are implemented in statistical software, though with a little work they can be obtained in most software packages
- There are some tests which technically could be applied in certain situations, but it is not very often seen (or recognized)
  - (I have denoted these cases with ?)

59

Technical Assumptions

- Approximate normal distribution for estimated parameters
  - Sufficiently large sample sizes
- Appropriate handling of correlated data
  - Independence vs matched samples
- Appropriate estimation of variability of estimated parameters
  - Homoscedasticity vs heteroscedasticity vs mean-variance relationship

60

Two Independent Samples

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
Entire Distn	Chi Sq	Chi Sq	Chi Sq	Kol-Sm	Modif Kol-Sm
Diff in Proportion	Chi Sq	Chi Sq	Chi Sq	Chi Sq	KM
Odds Ratio	Chi Sq; Fish Ex	Chi Sq; Fish Ex	Chi Sq; Fish Ex; Prop Odds	Chi Sq; Fish Ex	KM

Two Independent Samples

	Binary	Unordrd	Ordered		
		Nominal	Categ	Quant	Cens
Diff in Medians			?(Bstrap)	Bstrap	?(Bstrp)
Median Difference			?(Sign)	?(Sign)	
Ratio of Medians				?(AFT regr); ?(δ method)	AFT regr; ?(δ method)

Two Group Comparisons

- If one sample estimates approx normal, we can compute SE for each group individually, then use methods for combining estimates

For independent $\hat{\theta}_1 \sim N(\theta_1, se_1^2)$, $\hat{\theta}_2 \sim N(\theta_2, se_2^2)$

$$\hat{\theta}_1 + \hat{\theta}_2 \sim N(\theta_1 + \theta_2, se_1^2 + se_2^2)$$

$$\hat{\theta}_1 - \hat{\theta}_2 \sim N(\theta_1 - \theta_2, se_1^2 + se_2^2)$$

$$\hat{\theta}_1 / \hat{\theta}_2 \sim N\left(\frac{\theta_1}{\theta_2}, \frac{1}{\theta_2^2}\left(se_1^2 + \frac{\theta_1^2}{\theta_2^2} se_2^2\right)\right)$$

63

Two Independent Samples

	Binary	Unordrd	Ordered		
		Nominal	Categ	Quant	Cens
(Diff in) Arithmetic Means (of Diff)	Chi Sq		t test (eq,uneq vrnc)	t test (eq,uneq vrnc)	?(Restr Mean)
(Ratio of) Geometric Means (Ratio)				t test (eq,uneq vrnc) on logs	?(Restr Geom Mean)

Two Independent Samples

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
Hazard Ratio				Logrank	Logrank
Pr (Y > X)			Wilcox Rnk Sum	Wilcox Rnk Sum	Modif Wilcox
???			?(Wilcox Sgn Rnk)	?(Wilcox Sgn Rnk)	

Matched Comparisons

- Must account for covariance between estimates for each group
  - Covariance between estimates involves correlation and standard errors

For correlated $\hat{\theta}_1 \sim N(\theta_1, se_1^2)$; $\hat{\theta}_2 \sim N(\theta_2, se_2^2)$

$$\hat{\theta}_1 + \hat{\theta}_2 \sim N\left(\theta_1 + \theta_2, se_1^2 + se_2^2 + 2Cov\left(\hat{\theta}_1, \hat{\theta}_2\right)\right)$$
$$\hat{\theta}_1 - \hat{\theta}_2 \sim N\left(\theta_1 - \theta_2, se_1^2 + se_2^2 - 2Cov\left(\hat{\theta}_1, \hat{\theta}_2\right)\right)$$

66

Two Matched Samples

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
Entire Distn	McNemar (Sign)				
Diff in Proportion	McNemar (Sign)	McNemar (Sign)	McNemar (Sign)	McNemar (Sign)	
Odds Ratio	McNemar (Sign)	McNemar (Sign)	McNemar (Sign)	McNemar (Sign)	

Two Matched Samples

	Binary	Unordrd	Ordered		
		Nominal	Categ	Quant	Cens
Diff in Medians			?(Bstrap)	Bstrap	
Median Difference			Sign	Sign	
Ratio of Medians			?(Bstrap)	Bstrap	

Two Matched Samples

	Binary	Unordrd Nominal	Ordered Categ	Ordered Quant	Ordered Cens
(Diff in) Arithmetic Means (of Diff)	McNemar (Sign)		Paired t test	Paired t test	
(Ratio of) Geometric Means (Ratio)				Paired t test on logs	

Two Matched Samples

	Binary	Unordered Nominal	Ordered Categ	Ordered Quant	Ordered Cens
Hazard Ratio				(B518)	(B518)
Pr (Y > X)			Sign	Sign	
???			Wilcox Sgn Rnk	Wilcox Sgn Rnk	

Regression Methods

- We extend these methods to the case of the “infinite sample” problem
  - Borrowing information
  - Contrasts across multiple groups

Infinite Samples

- While we don’t really ever have (or care) about an infinite number of samples, it is easiest to use models that would allow that in order to handle
  - Continuous predictors of interest
  - Adjustment for other variables

Simple Regression

- General notation for simple regression model

$$g(\theta_i) = \beta_0 + \beta_1 \times X_i$$

$g(\ )$ "link" function used for modeling

β_0 "Intercept"

β_1 "Slope (for predictor X)"

- The link function is usually either none (means) or log (geom mean, odds, hazard)

73

Regression Models

- According to the parameter compared across groups
  - Means → Linear regression
  - Geom Means → Linear regression on logs of response
  - Odds → Logistic regression
  - Rates → Poisson regression
  - Hazards → Proportional Hazards regression
  - Quantiles → Parametric survival regression

74

Regression Models

- According to the contrast across groups
  - Difference (identity link function)
    - Means → Linear regression
  - Ratios (log link function)
    - Geom Means → Linear regression on logs of response
    - Odds → Logistic regression
    - Rates → Poisson regression
    - Hazards → Proportional Hazards regression
    - Quantiles → Parametric survival regression

75

Regression Methods

		Unordered	Ordered		
	Binary	Nominal	Categ	Quant	Cens
Entire Distn	Logist				
Diff in Prop	(Linear)	(Linear)	(Linear)	(Linear)	
Odds Ratio	Logist	Logist	Logist; Prop Odds	Logist	

Regression Methods

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
Diff in Medians					
Median Diffrence					
Ratio of Medians				Param Surv (AFT)	Param Surv (AFT)

Regression Methods

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
(Diff in) Arith Means (of Diff)	(Linear)		Linear	Linear	
(Ratio of) Geometric Means (Ratio)				Linear on logs	

Regression Methods

	Binary	Unordered	Ordered		
		Nominal	Categ	Quant	Cens
Hazard Ratio				Prop Hazard	Prop Hazard
Pr (Y > X)					
???					

"Everything is Regression"

- The most commonly used two sample tests are special cases of regression
- Regression with a binary predictor
  - Linear regression → t test
  - Logistic regression → chi square (score test)
  - Proportional hazards → logrank (score test)

Biost 518 Topics

- Multiple regression
  - Models, interpretation of parameters
  - Modeling associations (dose-response)
  - Interactions
  - Time varying covariates; clustered data
  - Prediction
  - Missing data
  - Diagnostics
  - Exploratory models

81